

# The analysis Summary

X Education encounters a high volume of leads but grapples with a meager lead conversion rate of approximately 30%. The organization's directive is to formulate a model that assigns a lead score to each prospect, augmenting the likelihood of conversion for individuals with higher scores. The CEO's aspiration is to achieve an ambitious lead conversion rate of around 80%.

## Data Cleaning :

- ✓ Columns with >40% null values were removed.
- ✓ For categorical columns, decisions on action were derived from value counts analysis, considering skewness, creating new categories, imputing high-frequency values, or discarding columns.
- ✓ Numerical categorical data was imputed with mode.
- ✓ Columns with only one unique response from customers were discarded.
- ✓ Additional data treatments included handling outliers, rectifying invalid entries, grouping low-frequency values, and mapping binary categorical values.

## Exploratory Data Analysis (EDA):

- ✓ Addressed data imbalance; only 38.5% leads converted.
- ✓ Univariate and bivariate analyses were conducted for categorical and numerical variables, revealing insights from 'Lead Origin', 'Current occupation', 'Lead Source', etc.
- ✓ Positive correlation was observed between time spent on the website and lead conversion.

## Data Preparation:

- ✓ Created encoded dummy features for categorical variables.
- ✓ Split data into 70:30 train and test sets.
- ✓ Applied feature scaling using standardization.
- ✓ Eliminated highly correlated columns to mitigate multicollinearity.
- ✓ Utilized Recursive Feature Elimination (RFE) and manual feature reduction to condense variables from 48 to 15.

## Model Building:

Developed a final model (logm4) with 12 variables through iterative elimination based on p-values < 0.05 and VIF < 5.

Evaluated three prior models before arriving at the stable Model 4.

Model performance was appraised via accuracy, sensitivity, specificity, and precision.

# The analysis Summary

## Model Evaluation:

Optimized cut-off point at 0.345 was selected through consideration of accuracy, sensitivity, and specificity.

Lead score assigned using the chosen cut-off, with metrics close to 80% for both train and test datasets.

## Recommendations:

Allocate additional budget for Welingak Website advertising to bolster visibility.

Introduce incentives/discounts for lead-referred conversions to encourage more referrals.

Intensify targeting of working professionals due to their higher conversion rate and potentially stronger financial capability.

In conclusion, the project recommends targeted actions to optimize lead conversion rates, based on comprehensive data analysis, model development, and evaluation.